

STATISTICAL INFERENCE ON TRANSFORMATION MODELS: A SELF-INDUCED SMOOTHING APPROACH

BY JUNYI ZHANG, ZHEZHEN JIN,
YONGZHAO SHAO AND ZHILIANG YING

Columbia University and New York University

This paper deals with a general class of transformation models that contains many important semiparametric regression models as special cases. It develops a self-induced smoothing for the maximum rank correlation estimator, resulting in simultaneous point and variance estimation. The self-induced smoothing does not require bandwidth selection, yet provides the right amount of smoothness so that the estimator is asymptotically normal with mean zero (unbiased) and variance-covariance matrix consistently estimated by the usual sandwich-type estimator. An iterative algorithm is given for the variance estimation and shown to numerically converge to a consistent limiting variance estimator. The approach is applied to a data set involving survival times of primary biliary cirrhosis patients. Simulations results are reported, showing that the new method performs well under a variety of scenarios.

1. Introduction. Consider the following class of regression models, with response variable denoted by Y and $(d+1)$ -dimensional covariate vector by \mathbf{X} ,

$$(1) \quad Y = H(\mathbf{X}'\boldsymbol{\beta} + \varepsilon)$$

where $\boldsymbol{\beta}$ is the unknown parameter vector, ε is the unobserved error term that is independent of \mathbf{X} with a completely unspecified distribution, and H is a monotone increasing, but otherwise unspecified function.

It is easily seen that this class of models contains many commonly used regression models as its submodels that are especially important in the econometrics and survival analysis literature. For example, with $H(u) = u$, (1) becomes the standard regression model with an unspecified error distribution; with $H(u) = u^\lambda$ ($\lambda > 0$), the Box-Cox transformation model (Box and Cox, 1964); with $H(u) = I[u \geq 0]$, the binary choice model (Maddala, 1983;

AMS 2000 subject classifications: 62G20, 65C60

Keywords and phrases: rank correlation, semiparametric models, self-induced smoothing, variance estimation, algorithmic convergence, asymptotic normality, U-process, Hoeffding's decomposition

McFadden, 1984); with $H(u) = uI[u \geq 0]$, a censored regression model (Tobin, 1958; Powell, 1984); with $H(u) = \exp(u)$, the accelerated failure times (AFT) model (Cox and Oakes, 1984; Kalbfleisch and Prentice, 2002); with ε having an extreme value density $f(w) = \exp(w - \exp(w))$, the Cox proportional hazards regression (Cox, 1972); with ε having the standard logistic distribution, the proportional odds regression (Bennett, 1983).

A basic tool for handling model (1) is the maximum rank correlation (MRC) estimator proposed in the econometrics literature by Han (1987). Because both the transformation function H and the error distribution are unspecified, not all components of β are identifiable. Without loss of generality, we shall assume henceforth that the last component, $\beta_{d+1} = 1$. Let $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$ be a random sample from (1). Han's MRC estimator, denoted by $\hat{\theta}_n$, is the maximizer of following objective function

$$(2) \quad Q_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} I[Y_i > Y_j] I[\mathbf{X}_i' \beta(\theta) > \mathbf{X}_j' \beta(\theta)],$$

where $I[\cdot]$ denotes the indicator function, \mathbf{X}' the transpose of \mathbf{X} , and θ the first d components of β , i.e. $\beta(\theta) = (\theta_1, \dots, \theta_d, 1)'$. Han (1987) proved that the MRC estimator $\hat{\theta}_n$ is strongly consistent under certain regularity conditions.

An important subsequent development is due to Sherman (1993), who made use of the empirical process theory and Hoeffding's decomposition to approximate the objective function, viewed as a U-process. He showed that $\hat{\theta}_n$ is, in fact, asymptotically normal under additional regularity conditions. Estimation of the transformation function H was studied by Chen (2002), who constructed a rank-based estimator and established its consistency and asymptotic normality.

In addition to the econometrics, model (1) also encompasses the main semiparametric models in survival analysis, where right censoring is a major feature. Under the right censorship, there is a censoring variable C and one observes $\tilde{Y} = Y \wedge C$ and $\Delta_i = I(Y_i \leq C_i)$. Khan and Tamer (2007) constructed the following partial rank correlation function as an extension of the rank correlation objective function (2),

$$(3) \quad Q_n^*(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \Delta_j I[\tilde{Y}_i > \tilde{Y}_j] I[\mathbf{X}_i' \beta(\theta) > \mathbf{X}_j' \beta(\theta)].$$

They showed that the resulting maximum partial rank correlation estimate (PRCE) $\hat{\theta}_n^*$, as the maximizer of $Q_n^*(\theta)$, is consistent and asymptotically normal.

Crucial for the statistical inference of (1) based on $\hat{\theta}_n$ is the consistent variance estimation. In standard objective (loss) function derived estimation, the asymptotic variance is usually estimated by a sandwich-type estimator of form $\hat{\mathbf{A}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{A}}^{-1}$ with $\hat{\mathbf{A}}$ being the second derivative of the objective function and $\hat{\mathbf{V}}$ an estimator of the variance of the first derivative (score). The challenge here, however, is that Q_n itself is a (discontinuous) step function that precludes automatic use of differentiation to obtain $\hat{\mathbf{A}}$. Furthermore, $\hat{\mathbf{V}}$ is also difficult to obtain since the score function cannot be derived directly from Q_n via differentiation. Sherman(1993) suggested using numerical derivatives of first and second orders to construct $\hat{\mathbf{A}}$ and $\hat{\mathbf{V}}$. His approach requires bandwidth selection for the derivative functions. It is unclear how stable the resulting variance estimator is. Alternatively, one may resort to bootstrap (Efron, 1979) or other resampling methods (e.g. Jin et al., 2001). These approaches require repeatedly solving the maximization of (2), which is discontinuous and often multidimensional when $d > 1$. The computational cost could therefore be prohibitive.

In this paper, we develop a self-induced smoothing method for rank correlation criterion function (2) so that the differentiation can be performed, while bypassing the bandwidth selection. Both point and variance estimators can be obtained simultaneously in a straightforward way that is typically used for smooth objective functions. The new method is motivated by a novel approach proposed in Brown and Wang (2005, 2007), where an elegant self-induced smoothing method was introduced for non-smooth estimating functions. Although our approach bears similarity with that of Brown and Wang (2005), it is far from clear why such self-induced smoothing is suitable for the discrete objective function (rank correlation). In fact, undersmoothing would make the Hessian (second derivative) unstable while oversmoothing would introduce significant bias. Through highly technical and tedious derivations, we will show that the proposed method does strike a right balance in terms of asymptotic unbiasedness and enough smoothness for differentiation (twice).

The rest of the paper is organized as follows. In Section 2, the new methods are described and related large sample properties are developed. In particular, we give construction for simultaneous point and variance estimation and show that the resulting point estimator is asymptotically normal and the variance estimator is consistent. In Section 3, the approach, along with the algorithm and large sample properties, is extended to handle survival data with right censoring. Simulation results are reported in Section 4, where application to a real data set is also given. Section 5 contains some concluding remarks. Additional technical proofs can be found in the Appendix.

2. Main Results. In this section we develop a self-induced smoothing method for the rank correlation criterion function defined by (2). It is divided into three subsections, with the first introducing the method and the algorithm, the second establishing large sample properties and the third covering proofs.

2.1. *Methods.* Since MRC estimator $\hat{\theta}_n$ is asymptotically normal (Sherman, 1993), its difference with the true parameter value, $\hat{\theta}_n - \theta$, should approximately be a Gaussian noise \mathbf{Z}/\sqrt{n} , where $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$ is a d -dimensional normal random vector with mean $\mathbf{0}$ and covariance matrix Σ . Assume that \mathbf{Z} is independent of data and let $E_{\mathbf{Z}}$ denote the expectation with respect to \mathbf{Z} given data. A self-induced smoothing for Q_n is $\tilde{Q}_n(\theta) = E_{\mathbf{Z}}Q_n(\theta + \mathbf{Z}/\sqrt{n})$. The self-induced smoothing using the limiting Gaussian distribution was originally proposed by Brown and Wang (2005) for certain non-smooth estimating functions.

To get an explicit form for \tilde{Q}_n , let Φ be the standard normal distribution function, $\mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j$, $\sigma_{ij} = \sqrt{(\mathbf{X}_{ij}^{(1)})' \Sigma \mathbf{X}_{ij}^{(1)}}$ where $\mathbf{X}_{ij}^{(1)}$ denotes the first d components of \mathbf{X}_{ij} . Then, it is easy to see that

$$(4) \quad \tilde{Q}_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} I[Y_i > Y_j] \Phi(\sqrt{n} \mathbf{X}_{ij}' \beta(\theta) / \sigma_{ij}).$$

We shall use $\tilde{\theta}_n = \arg \max_{\Theta} \tilde{Q}_n(\theta)$ to denote the corresponding estimator, which will be called the smoothed maximum rank correlation estimator (SMRCE). Here and in the sequel, Θ denotes the parameter space for θ .

REMARK 1. Smoothing is an appealing way for a simple solution to the inference problem associated with the MRCE. If \tilde{Q}_n were a usual smooth objective function, then its first derivative would become the score function and its second derivative could be used for variance estimation. Specifically, if we use \mathbf{V} to denote the limiting variance of the score scaled by n and \mathbf{A} the limit of the second derivative, then the asymptotic variance of the resulting estimator, scaled by n , should be of form $\mathbf{A}^{-1} \mathbf{V} \mathbf{A}^{-1}$. A consistent estimator could then be obtained by the plug-in method, i.e. replacing unknown parameters by their corresponding empirical estimators.

REMARK 2. It is unclear, however, whether or not the self-induced smooth will provide a right amount of smoothing, even in view of the results given in Brown and Wang (2005). With over-smoothing, $\tilde{\theta}_n$ may be asymptotically biased, i.e. the bias is not of order $o(n^{-1/2})$; with under-smoothing, the “score” function (first derivative of \tilde{Q}_n) may have multiple “spikes” and

thus the second derivative matrix (Hessian) of \tilde{Q}_n may not behave properly and certainly cannot be expected to provide a consistent variance estimator.

In Subsection 2.2, we show that the self-induced smoothing here does result in a right amount of smoothing in the sense that the bias is asymptotically negligible and the Hessian matrix behave properly. Before starting the theoretic developments, we first describe our method.

We first differentiate the smoothed objective function \tilde{Q}_n to get score

$$\tilde{\mathbf{S}}_n(\boldsymbol{\theta}) = \frac{1}{n(n-1)} \sum_{i < j} H_{ij} \phi \left(\frac{\sqrt{n} \mathbf{X}'_{ij} \boldsymbol{\beta}(\boldsymbol{\theta})}{\sigma_{ij}} \right) \frac{\sqrt{n} \mathbf{X}_{ij}^{(1)}}{\sigma_{ij}},$$

where $H_{ij} = \text{sgn}(Y_i - Y_j)$. This is a U-process of order 2 with kernel

$$\mathbf{s}_n(\mathbf{U}_i, \mathbf{U}_j) = \frac{1}{2} H_{ij} \phi \left(\frac{\sqrt{n} \mathbf{X}'_{ij} \boldsymbol{\beta}(\boldsymbol{\theta})}{\sigma_{ij}} \right) \frac{\sqrt{n} \mathbf{X}_{ij}^{(1)}}{\sigma_{ij}},$$

where \mathbf{U}_i denotes the pair (Y_i, \mathbf{X}_i) .

By Hoeffding's decomposition, the asymptotic variance of $\sqrt{n} \tilde{\mathbf{S}}_n(\boldsymbol{\theta})$ is approximated by

$$(5) \quad \hat{\mathbf{V}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{1}{n^3} \sum_{i=1}^n \left\{ \sum_j \left[H_{ij} \times \phi \left(\frac{\sqrt{n} \mathbf{X}'_{ij} \boldsymbol{\beta}}{\sigma_{ij}} \right) \frac{\sqrt{n} \mathbf{X}_{ij}^{(1)}}{\sigma_{ij}} \right] \right\}^{\otimes 2},$$

where, for a vector \mathbf{v} , $\mathbf{v}^{\otimes 2} = \mathbf{v} \mathbf{v}'$. Thus, $\hat{\mathbf{V}}_n(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma})$ is used to estimate \mathbf{V} , the middle part of the “sandwich” variance formula discussed in Remark 1.

As for \mathbf{A} , we differentiate $\tilde{\mathbf{S}}_n(\boldsymbol{\theta})$ to get

$$(6) \quad \hat{\mathbf{A}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{1}{2n(n-1)} \sum_{i \neq j} \left\{ H_{ij} \times \dot{\phi} \left(\frac{\sqrt{n} \mathbf{X}'_{ij} \boldsymbol{\beta}}{\sigma_{ij}} \right) \left[\frac{\sqrt{n} \mathbf{X}_{ij}^{(1)}}{\sigma_{ij}} \right]^{\otimes 2} \right\},$$

where $\dot{\phi}(z) = -z\phi(z)$ is the derivative of $\phi(z)$. Although the self-induced smoothing was motivated earlier with $\boldsymbol{\Sigma}$ being the limiting covariance matrix of the estimator, we will show later that for any positive definite matrix $\boldsymbol{\Sigma}$, $\hat{\mathbf{A}}_n(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma})$ converges to \mathbf{A} .

Note that the above discussions about \mathbf{A} and \mathbf{V} are not mathematically rigorous. This is because the kernel function for the score process is sample size n -dependent. The usual asymptotic theory for the U-process is not applicable. Indeed, our rigorous derivations, to be given in Subsection 2.3, are quite tedious, involving many approximations that are quite delicate.

Let

$$(7) \quad \hat{\mathbf{D}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \hat{\mathbf{A}}_n^{-1}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \times \hat{\mathbf{V}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \times \hat{\mathbf{A}}_n^{-1}(\boldsymbol{\theta}, \boldsymbol{\Sigma}).$$

If $\boldsymbol{\theta}$ is the true parameter value, then $\hat{\mathbf{D}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ converges to the limiting covariance matrix, which is the desired choice for $\boldsymbol{\Sigma}$ in the self-induced smoothing. Therefore, (7) leads to an iterative algorithm of form $\hat{\boldsymbol{\Sigma}}_n^{(k)} = \hat{\mathbf{D}}_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\Sigma}}_n^{(k-1)})$; see also Brown and Wang (2005). Specifically, we propose the following algorithm:

ALGORITHM 1. (SMRCE)

1. Compute the MRC estimator $\hat{\boldsymbol{\theta}}_n$ and set $\hat{\boldsymbol{\Sigma}}^{(0)}$ to be the identity matrix.
2. Update variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_n^{(k)} = \hat{\mathbf{D}}_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\Sigma}}_n^{(k-1)})$. Smooth the rank correlation $Q_n(\boldsymbol{\theta})$ using covariance matrix $\hat{\boldsymbol{\Sigma}}_n^{(k)}$. Maximize the resulting smoothed rank correlation to get an estimator $\hat{\boldsymbol{\theta}}_n^{(k)}$.
3. Repeat step 2 until $\hat{\boldsymbol{\theta}}_n^{(k)}$ converge.

2.2. Large-sample properties. This subsection is devoted to the large sample theory. The main results are: 1. the smoothed MRC estimator (SMRCE) is asymptotically equivalent to the MRC estimator; 2. the proposed method leads to a consistent variance estimator; and 3. the iterative algorithm for point and variance estimation converges numerically.

We first introduce notation as well as assumptions, which are similar to those in Sherman (1993) for the MRC estimator. Let

$$(8) \quad \tau(y, \mathbf{x}, \boldsymbol{\theta}) = E \left[I_{[y > Y]} I_{[(\mathbf{x} - \mathbf{X})' \boldsymbol{\beta}(\boldsymbol{\theta}) > 0]} + I_{[y < Y]} I_{[(\mathbf{x} - \mathbf{X})' \boldsymbol{\beta}(\boldsymbol{\theta}) < 0]} \right],$$

which is the projection of the kernel of U-process $Q_n(\boldsymbol{\theta})$. The expectation is taken for (\mathbf{X}, Y) . Also let

$$|\nabla_m| \tau(y, \mathbf{x}, \boldsymbol{\theta}) = \sum_{i_1, \dots, i_m} \left| \frac{\partial^m \tau(y, \mathbf{x}, \boldsymbol{\theta})}{\partial \theta_{i_1} \cdots \partial \theta_{i_m}} \right|.$$

The following Assumptions 1 and 2 are used in Han (1987) (see also Sherman, 1993) to establish consistency for the MRC estimator. For asymptotic normality, we need an additional regularity condition (Assumption 3) given in Sherman (1993).

ASSUMPTION 1. The true parameter value $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$, which is a compact subset of the d -dimensional Euclidean space \mathbb{R}^d .

ASSUMPTION 2. The support of \mathbf{X} is not contained in any linear subspace of \mathbb{R}^{d+1} . Conditional on the first d components of \mathbf{X} , the last component of \mathbf{X} has a density function with respect to the Lebesgue measure.

ASSUMPTION 3. There exists a neighborhood, \mathcal{N} , of $\boldsymbol{\theta}_0$ such that for each pair (y, \mathbf{x}) of possible values of (Y, \mathbf{X}) ,

- (i) The second derivatives of $\tau(y, \mathbf{x}; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ exist in \mathcal{N} .
- (ii) There is an integrable function $M_1(y, \mathbf{x})$ such that for all $\boldsymbol{\theta}$ in \mathcal{N} ,

$$\|\nabla_2 \tau(y, \mathbf{x}; \boldsymbol{\theta}) - \nabla_2 \tau(y, \mathbf{x}; \boldsymbol{\theta}_0)\|_2 \leq M_1(y, \mathbf{x})|\boldsymbol{\theta} - \boldsymbol{\theta}_0|.$$

- (iii) $E(|\nabla_1| \tau(Y, \mathbf{X}; \boldsymbol{\theta}_0))^2 < +\infty$.
- (iv) $E|\nabla_2| \tau(Y, \mathbf{X}; \boldsymbol{\theta}_0) < +\infty$.
- (v) The matrix $E\nabla_2 \tau(Y, \mathbf{X}; \boldsymbol{\theta}_0)$ is strictly negative definite.

PROPOSITION 1. (Sherman, 1993) Assume that Assumptions 1-3 hold. We have, uniformly over any $o_p(1)$ neighborhood of $\boldsymbol{\theta}_0$,

$$(9) \quad Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_0) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{A}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{\sqrt{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{W}_n + O_p(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^3) + o_p\left(\frac{1}{n}\right),$$

where $\mathbf{W}_n = \frac{1}{\sqrt{n}} \sum_i \nabla_1 \tau(Y_i, \mathbf{X}_i; \boldsymbol{\theta}_0)$, $2\mathbf{A}(\boldsymbol{\theta}) = E\nabla_2 \tau(Y, \mathbf{X}; \boldsymbol{\theta})$ and $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\theta}_0)$. Consequently, for the MRC estimator $\hat{\boldsymbol{\theta}}_n$,

$$(10) \quad \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{A}_0^{-1} \mathbf{W}_n + o_p(1) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}_0),$$

where $\mathbf{D}(\boldsymbol{\theta}) = \mathbf{A}^{-1}(\boldsymbol{\theta}) \mathbf{V}(\boldsymbol{\theta}) \mathbf{A}^{-1}(\boldsymbol{\theta})$, $\mathbf{V}(\boldsymbol{\theta}) = E(\nabla_1 \tau(Y, \mathbf{X}; \boldsymbol{\theta}) [\nabla_1 \tau(Y, \mathbf{X}; \boldsymbol{\theta})]')$ and $\mathbf{D}_0 = \mathbf{D}(\boldsymbol{\theta}_0)$.

Because of the standardization, the rank correlation criterion function Q_n is bounded by 1. It is not difficult to establish a uniform law of large numbers

$$(11) \quad \lim_n \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| = 0, \text{ a.s.},$$

where $Q(\boldsymbol{\theta})$ is the expectation of $Q_n(\boldsymbol{\theta})$; cf. Han (1987) and Sherman (1993). Likewise, we can show that such uniform convergence also holds for \tilde{Q}_n , i.e.

$$(12) \quad \lim_n \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\tilde{Q}_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| = 0, \text{ a.s.}$$

Note that the limit Q remains the same.

In the following theorem, we claim that the estimate obtained from maximizing the smoothed rank correlation function (4) is also asymptotically normal with the same asymptotic covariance matrix as Han's MRCE.

THEOREM 1. For any given positive definite matrix Σ , let $\tilde{Q}_n(\theta)$ be defined as in (4) and $\tilde{\theta}_n = \operatorname{argmax}_{\theta} E_Z \Gamma_n(\theta + Z/\sqrt{n}) \tilde{Q}_n(\theta)$. Then, under Assumptions 1-3, $\tilde{\theta}_n$ is consistent, $\tilde{\theta}_n \rightarrow \theta_0$ a.s. and asymptotically normal,

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}_0),$$

where \mathbf{D}_0 is defined as in Proposition 1. In addition, $\tilde{\theta}_n$ is asymptotically equivalent to $\hat{\theta}_n$ in the sense that $\tilde{\theta}_n = \hat{\theta}_n + o_p(n^{-1/2})$.

Recall that (7) defines the sandwich-type variance estimator by pretending that \tilde{Q}_n is a standard smooth objective function. Theorem 2 below shows that (7) is consistent.

THEOREM 2. Let $\hat{\theta}_n$ be the MRC estimator and $\hat{\mathbf{D}}_n$ be defined by (7). Then, for any fixed positive definite matrix Σ , $\hat{\mathbf{D}}_n(\hat{\theta}_n, \Sigma)$ converges in probability to \mathbf{D}_0 , the limiting variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

REMARK 3. The self-induced smoothing uses the limiting covariance matrix \mathbf{D}_0 as Σ . In practice, we may initially choose the identity matrix for Σ , which is the same way as the initial step in algorithm I. By Theorem 2.1, we know that the one-step estimator $\hat{\Sigma}_n^{(1)}$ in algorithm I converges in probability to the true covariance. However, this one-step estimator depends on the initial choice of Σ . Algorithm 1 is an iterative algorithm with the variance-covariance estimator converging to the fixed point of $\hat{\mathbf{D}}_n(\hat{\theta}_n, \Sigma) = \Sigma$.

Convergence of Algorithm 1 is ensured by the following theorem. For notational simplicity, we let $\operatorname{vech}(\mathbf{B})$ be the vectorization of matrix \mathbf{B} . For any function v of Σ ,

$$\left| \frac{\partial}{\partial \Sigma} \right| v = \sum_{i,j} \left| \frac{\partial}{\partial \Sigma_{i,j}} \right| v, \quad \frac{\partial v}{\partial \Sigma} = \left(\frac{\partial v}{\partial \Sigma_{1,1}}, \frac{\partial v}{\partial \Sigma_{2,1}}, \dots, \frac{\partial v}{\partial \Sigma_{d,d}} \right)',$$

where $\Sigma_{r,s}$ denotes the (r, s) entry of Σ .

THEOREM 3. Let $\hat{\Sigma}_n^{(k)}$ be defined as in Algorithm 1. Suppose that Assumptions 1-3 hold. Then there exist Σ_n^* , $n \geq 1$, such that for any $\epsilon > 0$, there exists N , such that for all $n > N$,

$$P\left(\lim_{k \rightarrow \infty} \hat{\Sigma}_n^{(k)} = \Sigma_n^*, \quad \|\Sigma_n^* - \mathbf{D}_0\| < \epsilon\right) > 1 - \epsilon.$$

REMARK 4. For a fixed n , Σ_n^* represents the fixed point matrix in the iterative algorithm. The above theorem shows that with probability approaching 1, the iterative algorithm converges to a limit, as $k \rightarrow \infty$, and the limit converges in probability to the limiting covariance matrix \mathbf{D}_0 .

REMARK 5. The speed of convergence of $\hat{\Sigma}_n^{(k)}$ to Σ_n^* is faster than any exponential rate in the sense that $\|\hat{\Sigma}_n^{(k)} - \Sigma_n^*\| = o(\eta^k)$ for any $\eta > 0$. This can be seen from Step 2 of Algorithm 1 in Subsection 2.1 and (13) below,

$$(13) \quad \sup_{\|\theta - \theta_0\| = o(1), \Sigma \in \mathcal{N}(\mathbf{D}_0)} \left| \frac{\partial}{\partial \Sigma} [\hat{\mathbf{D}}_n(\theta, \Sigma)]_{r,s} \right| = o_p(1),$$

which will be proved in the Appendix. Here $\mathcal{N}(\mathbf{D}_0)$ is a small neighborhood of \mathbf{D}_0 and Σ is a positive definite matrix.

2.3. *Proofs.* In this section, we provide proofs for (1) asymptotic equivalence of SMRCE to MRCE, (2) consistency of the induced variance estimator and (3) convergence of Algorithm 1. Some of the technical developments used in the proofs will be given in the Appendix.

PROOF OF THEOREM 1. Without loss of generality, we assume $\theta_0 = \mathbf{0}$. As in Subsection 2.1, let \mathbf{Z} be a d -variate normal random vector with mean $\mathbf{0}$ and covariance matrix Σ . Define

$$\tilde{Q}_n(\theta) = E_{\mathbf{Z}} Q_n(\theta + \mathbf{Z}/\sqrt{n}).$$

Let $\Gamma_n(\theta) = Q_n(\theta) - Q_n(\theta_0)$ and $\tilde{\Gamma}_n(\theta) = E_{\mathbf{Z}} \Gamma_n(\theta + \mathbf{Z}/\sqrt{n}) = \tilde{Q}_n(\theta) - Q_n(\theta_0)$. Define

$$\tilde{\theta}_n = \operatorname{argmax}_{\theta} [\tilde{Q}_n(\theta)] = \operatorname{argmax}_{\theta} \tilde{\Gamma}_n(\theta).$$

Let $\Omega_n = I[\|\mathbf{Z}\|_2 > 2d \log n]$, where $\|\mathbf{Z}\|_2 = \sqrt{\mathbf{Z}'\mathbf{Z}}$. Then $P(\Omega_n) = o(n^{-2})$ due to the Gaussian tail of \mathbf{Z} . Since $|Q_n(\theta)| \leq 1$ and $|\Gamma_n(\theta)| \leq 2$,

$$|E_{\mathbf{Z}} \{\Gamma_n(\theta + \mathbf{Z}/\sqrt{n}) I[\Omega_n]\}| \leq P(\Omega_n) = o(n^{-2}).$$

By the Cauchy-Schwarz inequality,

$$E_{\mathbf{Z}} \{|\mathbf{Z}| I[\Omega_n]\} = o(n^{-2}) \text{ and } E_{\mathbf{Z}} \{|\mathbf{Z}|^2 I[\Omega_n]\} = o(n^{-2}).$$

By (9), uniformly over $o(1)$ neighborhoods of $\mathbf{0}$,

$$\begin{aligned} E_{\mathbf{Z}} \{\Gamma_n(\theta + \mathbf{Z}/\sqrt{n}) I[\Omega_n^c]\} &= (1/2) E_{\mathbf{Z}} \{(\theta + \mathbf{Z}/\sqrt{n})' \mathbf{A}_0 (\theta + \mathbf{Z}/\sqrt{n}) I[\Omega_n^c]\} \\ &\quad + (1/\sqrt{n}) E_{\mathbf{Z}} \{(\theta + \mathbf{Z}/\sqrt{n})' \mathbf{W}_n I[\Omega_n^c]\} + o_p(E_{\mathbf{Z}} \{|\theta + \mathbf{Z}/\sqrt{n}|^2 I[\Omega_n^c]\} + \frac{1}{n}). \end{aligned}$$

Note that

$$E_{\mathbf{Z}}\{|\boldsymbol{\theta} + \mathbf{Z}/\sqrt{n}|^2 I[\boldsymbol{\Omega}_n^c]\} \leq 2(E_{\mathbf{Z}}|\boldsymbol{\theta}|^2 + E_{\mathbf{Z}}|\mathbf{Z}|^2/n) = O(|\boldsymbol{\theta}|^2 + 1/n).$$

Therefore, uniformly over $o(1)$ neighborhoods of $\mathbf{0}$, we have

$$(14) \quad \tilde{\Gamma}_n(\boldsymbol{\theta}) = (1/2)\boldsymbol{\theta}'\mathbf{A}_0\boldsymbol{\theta} + (1/\sqrt{n})\boldsymbol{\theta}'\mathbf{W}_n + E(\mathbf{Z}'\mathbf{A}_0\mathbf{Z})/2n + o_p(|\boldsymbol{\theta}|^2 + 1/n).$$

Replacing $\boldsymbol{\theta}$ in (14) with $\boldsymbol{\theta}_0 = \mathbf{0}$ and subtracting it from $\tilde{\Gamma}_n(\boldsymbol{\theta})$, we have

$$(15) \quad \tilde{\Gamma}_n(\boldsymbol{\theta}) - \tilde{\Gamma}_n(\boldsymbol{\theta}_0) = \frac{1}{2}\boldsymbol{\theta}'\mathbf{A}_0\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\boldsymbol{\theta}'\mathbf{W}_n + o_p(|\boldsymbol{\theta}|^2 + 1/n).$$

Combining (15) with Lemma 1 in the Appendix, we get,

$$(16) \quad \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{A}_0^{-1}\mathbf{W}_n + o_p(1).$$

Therefore, from (10) and (16), we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) = o_p(1).$$

Finally, strong consistency of $\tilde{\boldsymbol{\theta}}_n$ follows the uniform almost sure convergence of \tilde{Q}_n as stated in (12). This completes the proof. \square

PROOF OF THEOREM 2. For notational simplicity, we assume throughout the proof that $\boldsymbol{\Sigma}$ is the identity matrix. The same argument with modifications to include constants for up and lower bound may be applied to deal with a general covariance matrix $\boldsymbol{\Sigma}$.

We first show

$$(17) \quad \hat{\mathbf{A}}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \mathbf{A}(\boldsymbol{\theta}_0).$$

By definition, $[\hat{\mathbf{A}}_n(\boldsymbol{\theta})]_{r,s} = \partial^2 \tilde{Q}_n(\boldsymbol{\theta}) / (\partial \theta_r \partial \theta_s)$. As defined in (4), $\tilde{Q}_n(\boldsymbol{\theta})$ has the following integral representation,

$$\tilde{Q}_n(\boldsymbol{\theta}) = \int Q_n(\boldsymbol{\theta} + \mathbf{z}/\sqrt{n}) (2\pi)^{-\frac{d}{2}} \exp(-\frac{\|\mathbf{z}\|_2^2}{2}) d\mathbf{z}.$$

By change of variable $\mathbf{t} = \boldsymbol{\theta} + \mathbf{z}/\sqrt{n}$,

$$(18) \quad \tilde{Q}_n(\boldsymbol{\theta}) = \int Q_n(\mathbf{t}) K_n(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t},$$

where $K_n(\mathbf{t}, \boldsymbol{\theta}) = (2\pi)^{-\frac{d}{2}} n^{\frac{d}{2}} \exp(-\frac{n\|\mathbf{t} - \boldsymbol{\theta}\|_2^2}{2})$. From (18),

$$\frac{\partial}{\partial \theta_r} \tilde{Q}_n(\boldsymbol{\theta}) = \int Q_n(\mathbf{t}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}$$

and

$$\frac{\partial^2}{\partial \theta_r \partial \theta_s} \tilde{Q}_n(\boldsymbol{\theta}) = \int Q_n(\mathbf{t}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t},$$

where $\dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) = \partial K_n(\mathbf{t}, \boldsymbol{\theta}) / \partial \theta_r$ and $\ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) = \partial^2 K_n(\mathbf{t}, \boldsymbol{\theta}) / (\partial \theta_r \partial \theta_s)$.

In view of (6), to show (17), it suffices to prove

$$(19) \quad \int Q_n(\mathbf{t}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = [\mathbf{A}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1)$$

uniformly over $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = O(n^{-1/2})$. To show (19), we define

$$\boldsymbol{\Omega}_{n,r} = \left\{ \mathbf{t} : (t_r - \theta_r)^2 < \frac{4 \log n}{n}, \sum_{i \neq r} (t_i - \theta_i)^2 < \frac{2(d-1) \log n}{n} \right\}.$$

By Lemma 2(i) and the boundedness of $Q_n(\mathbf{t})$, we have,

$$\int_{(\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s})^c} Q_n(\mathbf{t}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = o(n^{-1/2}),$$

where \mathfrak{B}^c for set \mathfrak{B} denotes its complement. Therefore, (19) reduces to

$$(20) \quad \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} Q_n(\mathbf{t}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = [\mathbf{A}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1).$$

To show (20), we establish a quadratic expansion of $Q_n(\mathbf{t})$ for $\mathbf{t} \in \boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}$. Since $\|\mathbf{t} - \boldsymbol{\theta}\|_2 < \sqrt{4d \log n / n}$ for $\mathbf{t} \in \boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = O(n^{-1/2})$, it follows that $\|\mathbf{t} - \boldsymbol{\theta}_0\|_2 = o(1)$. Therefore, by (9),

$$(21) \quad \begin{aligned} Q_n(\mathbf{t}) &= Q_n(\boldsymbol{\theta}_0) + \frac{1}{2}(\mathbf{t} - \boldsymbol{\theta}_0)' \mathbf{A}(\boldsymbol{\theta}_0)(\mathbf{t} - \boldsymbol{\theta}_0) \\ &\quad + (\mathbf{t} - \boldsymbol{\theta}_0)' \mathbf{W}_n / \sqrt{n} + O_p(|\mathbf{t} - \boldsymbol{\theta}_0|^3) + o_p(1/n). \end{aligned}$$

Therefore, the left hand side of (20) equals **I** + **II** + **III** + **IV**, where

$$\begin{aligned} \text{I} &= \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} [O_p(|\mathbf{t} - \boldsymbol{\theta}_0|^3) + o_p(1/n)] \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}, \\ \text{II} &= Q_n(\boldsymbol{\theta}_0) \times \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}, \\ \text{III} &= \frac{\mathbf{W}'_n}{\sqrt{n}} \times \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} (\mathbf{t} - \boldsymbol{\theta}_0) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}, \\ \text{IV} &= \frac{1}{2} \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} (\mathbf{t} - \boldsymbol{\theta}_0)' \mathbf{A}(\boldsymbol{\theta}_0)(\mathbf{t} - \boldsymbol{\theta}_0) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}. \end{aligned}$$

By the definition of $\boldsymbol{\Omega}_{n,r}$,

$$|\mathbf{I}| \leq \left| O_p \left(\frac{(\log n)^{\frac{3}{2}}}{n\sqrt{n}} \right) + o_p(1/n) \right| \times \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} |\ddot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t}.$$

By Lemma 2(ii), $\mathbf{I} = o_p(1)$. Furthermore, $\mathbf{II} = o(n^{-1/2})$ due to Lemma 2(iii). Note that

$$\begin{aligned} \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} (\mathbf{t} - \boldsymbol{\theta}_0) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} &= \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} (\mathbf{t} - \boldsymbol{\theta}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \\ (22) \quad &+ (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \\ &= (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}, \end{aligned}$$

where the last equality follows from the fact that $\boldsymbol{\Omega}_{n,r}$ and $\boldsymbol{\Omega}_{n,s}$ are symmetric at $\boldsymbol{\theta}$ and $(\mathbf{t} - \boldsymbol{\theta}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta})$ is an odd function of $[\mathbf{t} - \boldsymbol{\theta}]_r$ for $r = 1, 2, \dots, d$. Combining this with Lemma 2(i), we have $\mathbf{III} = \mathbf{o}(n^{-1})$. Again by symmetry,

$$\begin{aligned} \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} (\mathbf{t} - \boldsymbol{\theta}_0)' \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{t} - \boldsymbol{\theta}_0) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \\ (23) \quad &= \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} (\mathbf{t} - \boldsymbol{\theta})' \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{t} - \boldsymbol{\theta}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \\ &+ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{A}(\boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t}. \end{aligned}$$

By Lemma 2 (i) and (iv), $\mathbf{IV} = [\mathbf{A}(\boldsymbol{\theta}_0)]_{r,s} + o(n^{-1/2})$. Combining the approximations for $\mathbf{I} - \mathbf{IV}$, we get (20).

Next we prove $\hat{\mathbf{V}}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \mathbf{V}(\boldsymbol{\theta}_0)$ by showing, componentwise,

$$(24) \quad [\hat{\mathbf{V}}_n(\boldsymbol{\theta})]_{r,s} = [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1)$$

uniformly over $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = O(n^{-1/2})$ for $r, s = 1, \dots, d$.

Define

$$q(\mathbf{u}, \tilde{\mathbf{u}}; \boldsymbol{\theta}) = I_{[y > \tilde{y}]} I_{[(\mathbf{x} - \tilde{\mathbf{x}})' \boldsymbol{\beta} > 0]} + I_{[y < \tilde{y}]} I_{[(\mathbf{x} - \tilde{\mathbf{x}})' \boldsymbol{\beta} < 0]},$$

where $\mathbf{u} = (y, \mathbf{x})$ and $\tilde{\mathbf{u}} = (\tilde{y}, \tilde{\mathbf{x}})$. In addition, let $\tau_n(\mathbf{u}, \boldsymbol{\theta}) = \int q(\mathbf{u}, \tilde{\mathbf{u}}; \boldsymbol{\theta}) \mathbb{F}_n(d\tilde{\mathbf{u}})$,

where $\mathbb{F}_n(\cdot)$ is the empirical distribution for \mathbf{u}_i 's. By definition,

$$[\hat{\mathbf{V}}_n(\boldsymbol{\theta})]_{r,s} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial}{\partial \theta_r} \int \tau_n(\mathbf{u}_i, \boldsymbol{\theta} + \frac{\mathbf{z}}{\sqrt{n}}) (2\pi)^{-\frac{d}{2}} e^{-\frac{\|\mathbf{z}\|_2^2}{2}} d\mathbf{z} \right] \\ \times \left[\frac{\partial}{\partial \theta_s} \int \tau_n(\mathbf{u}_i, \boldsymbol{\theta} + \frac{\tilde{\mathbf{z}}}{\sqrt{n}}) (2\pi)^{-\frac{d}{2}} e^{-\frac{\|\tilde{\mathbf{z}}\|_2^2}{2}} d\tilde{\mathbf{z}} \right].$$

Letting $\mathbf{t} = \boldsymbol{\theta} + \mathbf{z}/\sqrt{n}$ and $\boldsymbol{\omega} = \boldsymbol{\theta} + \tilde{\mathbf{z}}/\sqrt{n}$, we have

$$[\hat{\mathbf{V}}_n(\boldsymbol{\theta})]_{r,s} = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta_r} \int \tau_n(\mathbf{u}_i, \mathbf{t}) K_n(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \times \frac{\partial}{\partial \theta_s} \int \tau_n(\mathbf{u}_i, \boldsymbol{\omega}) K_n(\boldsymbol{\omega}, \boldsymbol{\theta}) d\boldsymbol{\omega} \\ = \frac{1}{n} \sum_{i=1}^n \int \tau_n(\mathbf{u}_i, \mathbf{t}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \times \int \tau_n(\mathbf{u}_i, \boldsymbol{\omega}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\boldsymbol{\omega} \\ = \int G_n(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega},$$

where $G_n(\mathbf{t}, \boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^n \tau_n(\mathbf{u}_i, \mathbf{t}) \tau_n(\mathbf{u}_i, \boldsymbol{\omega})$, which is bounded by 0 and 1. By Lemma 2 (vii),

$$(25) \quad [\hat{\mathbf{V}}_n(\boldsymbol{\theta})]_{r,s} = o(n^{-\frac{1}{2}}) \\ + \int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} G_n(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega}$$

uniformly over $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = O(n^{-\frac{1}{2}})$. Let $f(\mathbf{u}, \mathbf{v}, \mathbf{w}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = q(\mathbf{u}, \mathbf{v}; \boldsymbol{\theta}_1) \times q(\mathbf{u}, \mathbf{w}; \boldsymbol{\theta}_2)$ and $f^*(\mathbf{u}, \mathbf{v}, \mathbf{w}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, the symmetrized f . By definition,

$$(26) \quad G_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{1}{\binom{n}{3}} \sum_{i < j < k} f^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{u}_k; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ + \frac{1}{n} \times \frac{1}{\binom{n}{2}} \sum_{i < j} f^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{u}_j; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \triangleq U_n + \frac{1}{n} \tilde{U}_n.$$

Clearly U_n is a third-order U-statistics and \tilde{U}_n is a second-order U-statistics. Applying Hoeffding's decomposition (van der Vaart, 1998, section 12.3),

$$(27) \quad U_n = \sum_{c=0}^3 \binom{3}{c} U_{n,c},$$

where $U_{n,c}$ is a U-statistics of order c ($c = 0, 1, 2, 3$) and defined as

$$U_{n,c} = \frac{1}{\binom{3}{c}} \sum_{|B|=c} \frac{1}{\binom{n}{3}} \sum_i P_B [f^*(\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \mathbf{u}_{i_3}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)].$$

Here, adopting the notations from van der Vaart (1998, Section 11.4), we define $P_B[f^*(\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \mathbf{u}_{i_3}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$ as a projection of f^* such that

$$\begin{aligned} P_\emptyset f^* &= E f^*, \\ P_{\{i\}} f^* &= E[f^*|\mathbf{u}_i] - E f^*, \\ P_{\{i,j\}} f^* &= E[f^*|\mathbf{u}_i, \mathbf{u}_j] - E[f^*|\mathbf{u}_i] - E[f^*|\mathbf{u}_j] + E f^*, \\ P_{\{1,2,3\}} f^* &= E[f^*|\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] - \sum_{i \neq j} E[f^*|\mathbf{u}_i, \mathbf{u}_j] + \sum_{i=1,2,3} E[f^*|\mathbf{u}_i] - E f^*. \end{aligned}$$

We know from Hoeffding's decomposition that $U_{n,2}$ and $U_{n,3}$ are second- and third-order degenerated U-statistics with bounded kernels and thus of order $o_p(n^{-1})$ and $o_p(n^{-3/2})$; see Sherman (1992, Corollary 8). Therefore, by Lemma 2(vi),

$$(28) \quad \int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} U_{n,c} \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} = o_p(1), \text{ for } c = 2, 3.$$

Replacing $U_{n,c}$ by \tilde{U}_n/n in (28) also results in $o_p(1)$. Then combining this and (28) with (26) and (27), (25) reduces to

$$(29) \quad \begin{aligned} [\hat{\mathbf{V}}_n(\boldsymbol{\theta})]_{r,s} &= 3 \times \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} U_{n,1} \times \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} \\ &\quad + \int_{\boldsymbol{\Omega}_{n,r} \cap \boldsymbol{\Omega}_{n,s}} E f \times \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} + o_p(1). \end{aligned}$$

Let $f_1(\mathbf{u}_j; \mathbf{t}, \boldsymbol{\omega}) = E[f(\mathbf{u}, \mathbf{v}, \mathbf{w}; \mathbf{t}, \boldsymbol{\omega})|\mathbf{u} = \mathbf{u}_j]$, $f_2(\mathbf{v}_j; \mathbf{t}, \boldsymbol{\omega}) = E[f(\mathbf{u}, \mathbf{v}, \mathbf{w}; \mathbf{t}, \boldsymbol{\omega})|\mathbf{v} = \mathbf{v}_j]$ and $f_3(\mathbf{w}_j; \mathbf{t}, \boldsymbol{\omega}) = E[f(\mathbf{u}, \mathbf{v}, \mathbf{w}; \mathbf{t}, \boldsymbol{\omega})|\mathbf{w} = \mathbf{w}_j]$. We define $\tilde{G}_n(\mathbf{t}, \boldsymbol{\omega}) = \frac{1}{n} \sum_{j=1}^n f_1(\mathbf{u}_j; \mathbf{t}, \boldsymbol{\omega})$. By the definitions of $f(\mathbf{u}, \mathbf{v}, \mathbf{w}; \mathbf{t}, \boldsymbol{\omega})$ and $q(\mathbf{u}, \mathbf{v}; \boldsymbol{\theta})$, we

have $\tilde{G}_n(\mathbf{t}, \boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^n \tau(\mathbf{u}_i, \mathbf{t}) \tau(\mathbf{u}_i, \boldsymbol{\omega})$. By Lemma 3 and applying integration by parts twice,

$$\begin{aligned} \int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} \tilde{G}_n(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} &= o(n^{-1/2}) \\ &\quad + \int_{\tilde{\boldsymbol{\Omega}}_{n,r} \times \tilde{\boldsymbol{\Omega}}_{n,s}} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\partial \tau(\mathbf{u}_i, \boldsymbol{\theta} + \frac{\mathbf{z}}{\sqrt{n}})}{\partial \theta_r} \frac{\partial \tau(\mathbf{u}_i, \boldsymbol{\theta} + \frac{\tilde{\mathbf{z}}}{\sqrt{n}})}{\partial \theta_s} \right\} \prod_i d\Phi(z_i) d\Phi(\tilde{z}_i), \end{aligned}$$

where $\tilde{\boldsymbol{\Omega}}_{n,r} := \{\mathbf{z} : z_r^2 < 4 \log n, \sum_{i \neq r} z_i^2 < 2(d-1) \log n\}$. By Lemma 3,

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta_r} \tau(\mathbf{u}_i, \boldsymbol{\theta}_1^*) \frac{\partial}{\partial \theta_s} \tau(\mathbf{u}_i, \boldsymbol{\theta}_2^*) = [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1)$$

uniformly over $\{(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*) : \|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_0\|_2 = o(1), i = 1, 2\}$. Therefore,

$$\int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} \frac{1}{n} \sum_{j=1}^n f_1(\mathbf{u}_j; \mathbf{t}, \boldsymbol{\omega}) \times \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} = [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1).$$

Similarly, applying integration by parts and by Lemma 3 and 2(vi), we have

$$\int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} \frac{1}{n} \sum_{j=1}^n f_2(\mathbf{v}_j; \mathbf{t}, \boldsymbol{\omega}) \times \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} = [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1),$$

$$\int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} \frac{1}{n} \sum_{j=1}^n f_3(\mathbf{w}_j; \mathbf{t}, \boldsymbol{\omega}) \times \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} = [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1),$$

$$\int_{\boldsymbol{\Omega}_{n,r} \times \boldsymbol{\Omega}_{n,s}} E[f(\mathbf{u}, \mathbf{v}, \mathbf{w}; \mathbf{t}, \boldsymbol{\omega})] \times \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} = [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1).$$

Hence the right hand side of (29) is $[\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1)$, which gives (24).

From (17) and (24), $\hat{\mathbf{D}}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \mathbf{D}_0$. \square

PROOF OF THEOREM 3. From Theorem (2), we know that $\hat{\boldsymbol{\Sigma}}_n^{(1)} \xrightarrow{p} \mathbf{D}_0$ and $\hat{\boldsymbol{\Sigma}}_n(\hat{\boldsymbol{\theta}}_n, \mathbf{D}_0) \xrightarrow{p} \mathbf{D}_0$. By the mean value theorem,

$$\begin{aligned} [\hat{\boldsymbol{\Sigma}}_n^{(2)} - \mathbf{D}_0]_{r,s} &= [\hat{\boldsymbol{\Sigma}}_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\Sigma}}_n^{(1)}) - \hat{\boldsymbol{\Sigma}}_n(\hat{\boldsymbol{\theta}}_n, \mathbf{D}_0)]_{r,s} + [\hat{\boldsymbol{\Sigma}}_n(\hat{\boldsymbol{\theta}}_n, \mathbf{D}_0) - \mathbf{D}_0]_{r,s} \\ &= \left[\frac{\partial}{\partial \boldsymbol{\Sigma}} [\hat{\mathbf{D}}_n]_{r,s} \Big|_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}^*} \right]' \times \text{vech}(\hat{\boldsymbol{\Sigma}}_n^{(1)} - \mathbf{D}_0) + [\hat{\boldsymbol{\Sigma}}_n(\hat{\boldsymbol{\theta}}_n, \mathbf{D}_0) - \mathbf{D}_0]_{r,s}, \end{aligned}$$

where $\|\boldsymbol{\Sigma}^* - \mathbf{D}_0\| \leq \|\hat{\boldsymbol{\Sigma}}_n^{(1)} - \mathbf{D}_0\|$ and thus $\boldsymbol{\Sigma}^* \in \mathcal{N}(\mathbf{D}_0)$. In view of Lemma 4 and $\hat{\boldsymbol{\Sigma}}_n(\hat{\boldsymbol{\theta}}_n, \mathbf{D}_0) \xrightarrow{p} \mathbf{D}_0$, $\hat{\boldsymbol{\Sigma}}_n^{(2)} \xrightarrow{p} \mathbf{D}_0$. Again by the mean value theorem,

$$[\hat{\boldsymbol{\Sigma}}_n^{(k+1)} - \hat{\boldsymbol{\Sigma}}_n^{(1)}]_{r,s} = \left[\frac{\partial}{\partial \boldsymbol{\Sigma}} [\hat{\mathbf{D}}_n]_{r,s} \Big|_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}^*} \right]' \times \text{vech}(\hat{\boldsymbol{\Sigma}}_n^{(k)} - \mathbf{D}_0),$$

where $\|\boldsymbol{\Sigma}^* - \mathbf{D}_0\| \leq \|\hat{\boldsymbol{\Sigma}}_n^{(k)} - \mathbf{D}_0\|$. Then by Lemma 4 and mathematical induction, we know that for any $\epsilon > 0$ and $\eta > 0$, there exist K and N , such that for any $n > N$ and $k > K$,

$$P\left(|[\hat{\boldsymbol{\Sigma}}_n^{(k+1)} - \hat{\boldsymbol{\Sigma}}_n^{(k)}]_{r,s}| \leq \eta \times |[\hat{\boldsymbol{\Sigma}}_n^{(k)} - \hat{\boldsymbol{\Sigma}}_n^{(k-1)}]_{r,s}|, \text{ for all } k > K\right) > 1 - \epsilon,$$

where $1 \leq s, r \leq d$. Note that the inequality inside the above probability implies that $\hat{\Sigma}_n^{(k)}$ converges as $k \rightarrow \infty$ and the limit Σ_n^* satisfies $\Sigma_n^* = \hat{\mathbf{D}}_n(\hat{\boldsymbol{\theta}}_n, \Sigma_n^*)$ and $\Sigma_n^* \xrightarrow{p} \mathbf{D}_0$. \square

3. Extensions. In this section, we extend the approach to the partial rank correlation (PRC) criterion function Q_n^* , defined by (3), of Khan and Tamer (2007) for censored data. Under the usual conditional independence between failure and censoring times given covariates and additional regularity conditions, Khan and Tamer (2007) developed asymptotic properties for PRCE that are parallel to those by Sherman (1993).

The same self-induced smoothing can be applied to partial rank correlation criteria function to get

$$(29) \quad \begin{aligned} \tilde{Q}_n^*(\boldsymbol{\theta}) &= E_{\mathbf{Z}} Q_n^*(\boldsymbol{\theta} + \mathbf{Z}/\sqrt{n}) \\ &= \frac{1}{n(n-1)} \sum_{i \neq j} \Delta_j I[\tilde{Y}_i > \tilde{Y}_j] \Phi(\sqrt{n} \mathbf{X}_{ij}' \boldsymbol{\beta}(\boldsymbol{\theta}) / \sigma_{ij}). \end{aligned}$$

We define its maximizer, $\tilde{\boldsymbol{\theta}}_n^*$, as the smoothed partial rank correlation estimator (SPRCE). Let

$$(30) \quad \hat{\mathbf{A}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{1}{2n(n-1)} \sum_{i \neq j} \left\{ H_{ij} \times \phi\left(\frac{\sqrt{n} \mathbf{X}_{ij}' \boldsymbol{\beta}}{\sigma_{ij}}\right) \left[\frac{\sqrt{n} \mathbf{X}_{ij}^{(1)}}{\sigma_{ij}}\right]^{\otimes 2} \right\},$$

$$(31) \quad \hat{\mathbf{V}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{1}{n^3} \sum_{i=1}^n \left\{ \sum_j \left[H_{ij} \times \phi\left(\frac{\sqrt{n} \mathbf{X}_{ij}' \boldsymbol{\beta}}{\sigma_{ij}}\right) \frac{\sqrt{n} \mathbf{X}_{ij}^{(1)}}{\sigma_{ij}} \right] \right\}^{\otimes 2},$$

$$(32) \quad \hat{\mathbf{D}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = [\hat{\mathbf{A}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma})]^{-1} \times \hat{\mathbf{V}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \times [\hat{\mathbf{A}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma})]^{-1},$$

where $H_{ij} = \Delta_j \times I[\tilde{Y}_i > \tilde{Y}_j] - \Delta_i \times I[\tilde{Y}_j > \tilde{Y}_i]$.

Based on $\hat{\mathbf{D}}_n^*(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, we have the following iterative algorithm to compute the SPRCE and variance estimate simultaneously.

ALGORITHM 2. (SPRCE)

1. Compute the PRC estimator $\hat{\boldsymbol{\theta}}_n^*$ and set $\hat{\boldsymbol{\Sigma}}^{(0)}$ to be the identity matrix.

2. Update variance-covariance matrix $\hat{\Sigma}_n^{*(k)} = \hat{\mathbf{D}}_n^*(\hat{\theta}_n^*, \hat{\Sigma}_n^{*(k-1)})$. Smooth the partial rank correlation $Q_n^*(\theta)$ using covariance matrix $\hat{\Sigma}_n^{*(k)}$. Maximize the resulting smoothed partial rank correlation to get an estimator $\hat{\theta}_n^{*(k)}$.
3. Repeat step 2 until $\hat{\theta}_n^{*(k)}$ converge.

In addition to Assumptions 1-3, Khan and Tamer (2007) added the following assumption for the consistency of PRCE.

ASSUMPTION 4. Let \mathbf{S}_X be the support of \mathbf{X}_i , and \mathfrak{X}_{uc} be the set

$$\mathfrak{X}_{uc} = \{\mathbf{x} \in \mathbf{S}_X : P(\Delta_i = 1 | \mathbf{X}_i = \mathbf{x}) > 0\}.$$

Then $P(\mathfrak{X}_{uc}) > 0$.

Similar to the rank correlation function, it can be shown that under Assumptions 1-4, (9) and (11) still hold for partial rank correlation function $Q_n^*(\theta)$. Therefore, Theorems 1-3 in Section 2 continue to hold when replacing the point and variance estimators for smoothed rank correlation by the corresponding ones for the smoothed partial rank correlation. Specifically, for any positive definite matrix Σ , under Assumptions 1-4, we have

1. The SPRCE $\tilde{\theta}_n^*$ is asymptotically equivalent to the PRCE $\hat{\theta}_n^*$ in the sense that $\tilde{\theta}_n^* = \hat{\theta}_n^* + o_p(n^{-1/2})$, and, therefore,

$$\sqrt{n}(\tilde{\theta}_n^* - \theta_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}_0^*),$$

where \mathbf{D}_0^* is the limiting variance-covariance matrix of $\hat{\theta}_n^*$.

2. Variance estimator is consistent: $\hat{\mathbf{D}}_n^*(\hat{\theta}_n^*, \Sigma) \xrightarrow{p} \mathbf{D}_0^*$.
3. Algorithm 2 converges numerically in the sense that there exist Σ_n^* , $n \geq 1$, such that for any $\epsilon > 0$, there exists N , such that for all $n > N$, $P(\lim_{k \rightarrow \infty} \hat{\Sigma}_n^{*(k)} = \Sigma_n^*, \|\Sigma_n^* - \mathbf{D}_0^*\| < \epsilon) > 1 - \epsilon$.

The proofs are similar to those of Theorems 1-3 in Section 2, and are, therefore, omitted.

4. Numerical results. In this section, we first apply the proposed self-induced smoothing method to analyze the primary biliary cirrhosis (PBC) data (Fleming and Harrington, 1990, Appendix D) and compare the result with that using the Cox regression. We then report results from several simulation studies we conducted using the method.

4.1. *PBC data.* We applied smoothed PRCE to the survival times of the first 312 subjects with no missing covariates in the PBC data. We included two covariates albumin and age50 (age divided by 50). We reparameterized the transformation model (1) by setting β_{age50} as 1, and estimated $\theta_{albumin}$ by SPRCE. We also calculated PRCE for $\theta_{albumin}$ and fitted the standard Cox model. For the Cox regression, the ratio $\hat{\beta}_{albumin}/\hat{\beta}_{age50}$ is the estimate of $\theta_{albumin}$. The results are summarized in Table 1.

Table 1
Regression Analysis of PBC data

	Albumin	SE
SPRCE	-4.29	1.40
PRCE	-3.50	-
Cox	-3.04	0.60

Note that PRCE does not have a readily available standard error estimate. The standard error of $\hat{\beta}_{albumin}/\hat{\beta}_{age50}$ in the Cox model was estimated by the delta method. Estimates from both the SPRCE and the Cox model conclude that the ratio of $\beta_{albumin}$ to β_{age50} is significant.

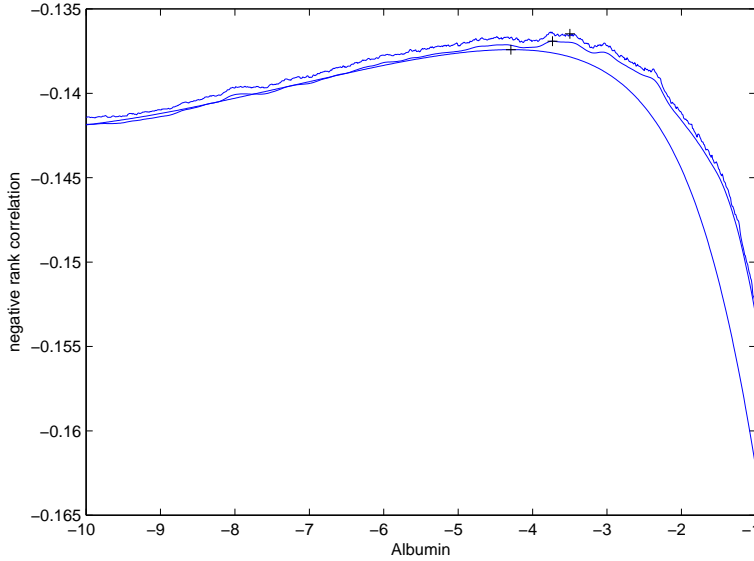


FIG 1.

To further assess the self-induced smoothing procedure, we plot the original objective function as well as the smoothed one in the first and last steps of our algorithm, as shown in Figure 1. The top curve is the original objective function, the middle curve is one after the initial smoothing, and the bottom curve is the limit of the iterative algorithm (after 8 iterations). It appears that the one-step smoothed objective function is under-smoothed in terms of the level of fluctuations, and the limiting curve is quite smooth.

Table 2
The proportional hazard model without censoring

$n = 500$	Est	Mean	Bias	RMSE	SE	coverage
θ	SMRCE	1.601	1.2×10^{-3}	0.0298	0.0316	92.3%
	MRCE	1.601	0.8×10^{-3}	0.0340	-	-
	Cox	1.599	-0.9×10^{-3}	0.0200	-	-
$n = 1000$	Est	Mean	Bias	RMSE	SE	coverage
θ	SMRCE	1.601	1.0×10^{-3}	0.0193	0.0212	93.9%
	MRCE	1.600	0.2×10^{-3}	0.0225	-	-
	Cox	1.600	0.1×10^{-3}	0.0141	-	-
$n = 2000$	Est	Mean	Bias	RMSE	SE	coverage
θ	SMRCE	1.600	0.2×10^{-3}	0.0136	0.0144	94.9%
	MRCE	1.600	-0.1×10^{-3}	0.0158	-	-
	Cox	1.600	0.1×10^{-3}	0.0100	-	-

Table 3
The proportional hazard model with censoring

$n = 600$	Est	Mean	Bias	RMSE	SE	coverage
θ	SPRCE	1.604	3.7×10^{-3}	0.0282	0.0300	93.2%
	PRCE	1.603	2.9×10^{-3}	0.0327	-	-
	Cox	1.601	1.0×10^{-3}	0.0204	-	-
$n = 1200$	Est	Mean	Bias	RMSE	SE	coverage
θ	SPRCE	1.601	1.1×10^{-3}	0.0190	0.0201	93.9%
	PRCE	1.601	0.8×10^{-3}	0.0217	-	-
	Cox	1.600	-0.2×10^{-3}	0.0139	-	-
$n = 2400$	Est	Mean	Bias	RMSE	SE	coverage
θ	SPRCE	1.600	0.4×10^{-3}	0.0127	0.0136	95.4%
	PRCE	1.600	0.1×10^{-3}	0.0148	-	-
	Cox	1.600	-0.2×10^{-3}	0.0097	-	-

Table 4
The linear model with gaussian noise

$n = 250$	Est	Mean	Bias	RMSE	SE	coverage
θ_1	SMRCE	1.615	1.5×10^{-2}	0.0747	0.0756	91.7%
	MRCE	1.612	1.2×10^{-2}	0.0730	-	-
	LS	1.601	0.7×10^{-3}	0.0296	-	-
θ_2	SMRCE	.5042	0.4×10^{-2}	0.0427	0.0443	93.6%
	MRCE	.5058	0.5×10^{-2}	0.0423	-	-
	LS	.5006	0.6×10^{-3}	0.0354	-	-
$n = 500$	Est	Mean	Bias	RMSE	SE	coverage
θ_1	SMRCE	1.605	4.9×10^{-3}	0.0515	0.0513	92.7%
	MRCE	1.607	6.7×10^{-3}	0.0523	-	-
	LS	1.601	0.7×10^{-3}	0.021	-	-
θ_2	SMRCE	.5023	2.3×10^{-3}	0.0296	0.0302	94.6%
	MRCE	.5042	4.2×10^{-3}	0.0316	-	-
	LS	.5006	0.6×10^{-3}	0.0254	-	-
$n = 1000$	Est	Mean	Bias	RMSE	SE	coverage
θ_1	SMRCE	1.603	3.6×10^{-3}	0.0361	0.0348	92.4%
	MRCE	1.603	3.4×10^{-3}	0.0382	-	-
	LS	1.601	0.5×10^{-3}	0.0144	-	-
θ_2	SMRCE	.5009	0.9×10^{-3}	0.0203	0.0207	94.8%
	MRCE	.5018	1.8×10^{-3}	0.0214	-	-
	LS	.5004	0.4×10^{-3}	0.0176	-	-

4.2. *Simulation studies.* We conducted simulation studies for a number of cases. In the first case (Design I), we generated \mathbf{X} from a bivariate normal distribution with mean $[-10, 20]'$ and a covariance matrix $\text{diag}\{3^2, 2^2\}$. We set $\beta_0^T = (\theta, 1) = [1.6, 1]$ and generated ϵ from the probability density function $f(w) = 2\exp(2w - \exp(2w))$. We set the transformation $H(x)$ as $H^{-1}(y) = \log(y^2)$. This is indeed a Weibull proportional hazard model. The sample sizes were $n = 500, 1000, 2000$ and the numbers of replications were 500. The SMRCE, MRCE and Cox model were used to estimate θ , and the standard error of SMRCE was computed by Algorithm 1. The mean(Mean), bias(Bias) and root mean square error(RMSE) for each method as well as mean of standard error(SE) and coverage of 95% confidence interval for the SMRCE are reported in Table 2.

The second case (Design II) is similar to the first one except that Y is censored by a random variable C , which is independent of \mathbf{X} and normally distributed with mean $\mu = 9.2$ and variance $\sigma^2 = 0.5^2$. The sample sizes were

$n = 600, 1200, 2400$ and the numbers of replications were 500. This design is similar to that in Gørgens and Horowitz (1999). The SPRCE, PRCE and Cox model were used to estimate θ , and the standard error of SPRCE was computed by Algorithm 2. The resulting estimates are summarized in Table 3 where we also report bias(Bias), root mean square error(RMSE), mean of standard error(SE), and coverage of 95% confidence interval.

In the third case (Design III), we generated $\mathbf{X} = [X_1, X_2, X_3]'$ by two steps. We first generated $[X_1, X_3]'$ from a bivariate normal distribution with mean $[-2, 2]'$ and an identity covariance matrix. We then generated X_2 as 0 or 2 with equal probability. We set $\beta_0^T = (\theta_1, \theta_2, 1) = [1.6, 0.5, 1]$ and generated ϵ from a normal distribution with $\mu = 0$ and $\sigma^2 = 0.5^2$. We set the transformation $H(x) = x$. The sample sizes were $n = 250, 500, 1000$ and the numbers of replications were 500. The SMRCE, MRCE and least squared method were applied to estimate θ_1 and θ_2 , and the standard error of SMRCE was computed by Algorithm 1. Table 4 reports the mean (Mean), bias (Bias) and root mean square error (RMSE) for each method as well as mean of standard error (SE) and coverage of 95% confidence interval for the SMRCE.

From Tables 2, 3 and 4, we find that (1) the root mean squared error is close to the mean standard error for the SMRCE (SPRCE); (2) as the sample size increases, the bias reduces and the coverage of 95% confidence interval converges to the nominal level. These show that the proposed variance estimator is accurate and Algorithms 1 and 2 work well.

5. Discussion. This paper provides a simple yet general recipe for smoothing the discontinuous rank correlation criteria function. The smoothing is self-induced in the sense that the implied bandwidth is essentially the asymptotic standard deviation of the regression parameter estimator. It is shown that such smoothing does not introduce any significant bias in that the resulting estimator is asymptotically equivalent to the original maximum rank correlation estimator, which is asymptotically normal. The smoothed rank correlation can be used as if it were a regular smooth criterion function in the usual M-estimation problem, in the sense that the standard sandwich-type plug-in variance-covariance estimator is consistent. Simulation and real data analysis provide additional evidence that the proposed method gives the right amount of smoothing.

Because of the family of transformation models contains both the proportional hazards and accelerated failure time models as its submodels, the new approach may be used for model selection. The specification test commonly used in the econometrics literature (Hausman, 1978) may also be

used for testing a specific semiparametric model assumption. In addition, the smoothed objective function also makes it possible to fit a penalized regression by introducing a LASSO-type penalty.

The method and theory developed herein can easily be extended to other problems of similar nature, i.e. discontinuous objective functions with associated estimators being asymptotically normal. In particular, we can apply the self-induced smoothing to the estimator introduced by Chen (2002) for the transformation function H to obtain a consistent variance estimator.

APPENDIX A: LEMMAS, COROLLARIES AND PROOFS

Lemma 1 below is due to Sherman (1993, Theorem 2).

LEMMA 1. *We denote $\Gamma_n(\boldsymbol{\theta})$ as general objective functions which are centered and satisfies the same regularity conditions as in Sherman (1993). Suppose $\boldsymbol{\theta}_n := \operatorname{argmax}_{\boldsymbol{\Theta}} \Gamma_n(\boldsymbol{\theta})$ is consistent for $\boldsymbol{\theta}_0$, an interior point of $\boldsymbol{\Theta}$. Suppose also that uniformly over $o_p(1)$ neighborhoods of $\boldsymbol{\theta}_0$,*

$$\Gamma_n(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{\sqrt{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{W}_n + o_p(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2) + o_p(1/n)$$

where \mathbf{A} is a negative definite matrix, and \mathbf{W}_n converges in distribution to a $N(\mathbf{0}, \mathbf{V})$ random vector. Then

$$\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = -\mathbf{A}^{-1} \mathbf{W}_n + o_p(1) \xrightarrow{L} N(\mathbf{0}, \mathbf{A}^{-1} \mathbf{V} \mathbf{A}^{-1}).$$

Recall in Theorem 2, we define $K_n(\mathbf{t}, \boldsymbol{\theta}) = (2\pi)^{-\frac{d}{2}} n^{\frac{d}{2}} \exp(-\frac{n\|\mathbf{t} - \boldsymbol{\theta}\|_2^2}{2})$ and its first and second partial derivatives with respect to $\boldsymbol{\theta}$ as

$$\begin{aligned} \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) &= \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} n(t_r - \theta_r) e^{-\frac{n\|\mathbf{t} - \boldsymbol{\theta}\|_2^2}{2}}, \\ \ddot{K}_{n,r,r}(\mathbf{t}, \boldsymbol{\theta}) &= \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} n(n(t_r - \theta_r)^2 - 1) e^{-\frac{n\|\mathbf{t} - \boldsymbol{\theta}\|_2^2}{2}}, \\ \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) &= \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} n^2(t_r - \theta_r)(t_s - \theta_s) e^{-\frac{n\|\mathbf{t} - \boldsymbol{\theta}\|_2^2}{2}}. \end{aligned}$$

Also recall

$$\boldsymbol{\Omega}_{n,r} = \{\mathbf{t} : (t_r - \theta_r)^2 < 4 \log n/n, \sum_{i \neq r} (t_i - \theta_i)^2 < 2(d-1) \log n/n\}.$$

Then we have the following lemma.

LEMMA 2. *Uniformly over $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = O(n^{-\frac{1}{2}})$,*

- (i) $\int_{(\Omega_{n,r} \cap \Omega_{n,s})^c} F(\mathbf{t}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = o(\frac{1}{\sqrt{n}}), \forall F(\mathbf{t}) \text{ s.t. } 0 \leq F(\mathbf{t}) \leq 1.$
- (ii) $\int_{\Omega_{n,r} \cap \Omega_{n,s}} \frac{1}{n} |\ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} = O(1).$
- (iii) $\int_{\Omega_{n,r} \cap \Omega_{n,s}} \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = o(n^{-1/2}).$
- (iv) $\int_{\Omega_{n,r} \cap \Omega_{n,s}} \frac{1}{2} (\mathbf{t} - \boldsymbol{\theta})' \mathbf{A} (\mathbf{t} - \boldsymbol{\theta}) \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = [A]_{r,s} + o(n^{-1}).$
- (v) $\int_{\Omega_{n,r}^c} |\dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} = O(n^{-3/2}).$
- (vi) $\int_{\Omega_{n,r}} \frac{1}{\sqrt{n}} |\dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} = O(1).$
- (vii) For any given $0 \leq G(\mathbf{t}, \boldsymbol{\omega}) \leq 1,$

$$\begin{aligned} & \int G(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} \\ &= \int_{\Omega_{n,r} \times \Omega_{n,s}} G(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} + o(n^{-\frac{1}{2}}). \end{aligned}$$

PROOF. Let $\tilde{\Omega}_{n,r} = \{\mathbf{t} : t_r^2 < 4 \log n/n, \sum_{i \neq r} t_i^2 < 2(d-1) \log n/n\}$, and divide its complement into $\tilde{\Omega}_{n,r}^{(1)} := \{\mathbf{t} : t_r^2 > 4 \log n/n\}$ and $\tilde{\Omega}_{n,r}^{(2)} := \{\mathbf{t} : t_r^2 < 4 \log n/n, \sum_{i \neq r} t_i^2 \geq 2(d-1) \log n/n\}$. We prove (i)-(iv) for $s = r$. For $s \neq r$, the proofs are similar and omitted.

For (i), note that

$$\int_{\Omega_{n,r}^c} F(\mathbf{t}) \ddot{K}_{n,r,r}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} = \int_{\tilde{\Omega}_{n,r}^c} F(\mathbf{t} + \boldsymbol{\theta}) \ddot{K}_{n,r,r}(\mathbf{t}, \mathbf{0}) d\mathbf{t}.$$

Since $0 \leq F(\mathbf{t}) \leq 1$ and $(nt_r^2 - 1)I[\tilde{\Omega}_{n,r}^{(1)}] \geq 0$,

$$\begin{aligned} \int_{\tilde{\Omega}_{n,r}^{(1)}} F(\mathbf{t} + \boldsymbol{\theta}) \ddot{K}_{n,r,r}(\mathbf{t}, \mathbf{0}) d\mathbf{t} &= \int_{\tilde{\Omega}_{n,r}^{(1)}} F(\mathbf{t} + \boldsymbol{\theta}) \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} n(nt_r^2 - 1) e^{-\frac{n\|\mathbf{t}\|_2^2}{2}} d\mathbf{t} \\ &\leq \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} \int_{\tilde{\Omega}_{n,r}^{(1)}} n(nt_r^2 - 1) e^{-\frac{n\|\mathbf{t}\|_2^2}{2}} d\mathbf{t} \\ &= (2\pi)^{-\frac{1}{2}} \int_{\tilde{\Omega}_{n,r}^{(1)}} n^{\frac{1}{2}} d(nt_r e^{-\frac{nt_r^2}{2}}) \prod_{i \neq r} d\Phi(\sqrt{n}t_i) = o\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

where the last equality follows from $0 \leq \int \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq 1$. Similarly,

$$\begin{aligned} \int_{\tilde{\Omega}_{n,r}^{(2)}} F(\mathbf{t} + \boldsymbol{\theta}) \ddot{K}_{n,r,r}(\mathbf{t}, \mathbf{0}) d\mathbf{t} &= \int_{\tilde{\Omega}_{n,r}^{(2)}} F(\mathbf{t} + \boldsymbol{\theta}) \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} (n^2 t_r^2 - n) e^{-\frac{n\|\mathbf{t}\|_2^2}{2}} d\mathbf{t} \\ &\leq (2\pi)^{-\frac{1}{2}} \int_{\tilde{\Omega}_{n,r}^{(2)}} n|nt_r^2 - 1| e^{-\frac{nt_r^2}{2}} d\sqrt{n}t_r \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq 8 \frac{n(\log n)^{3/2}}{n^2} = o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

For (ii), by definition,

$$\begin{aligned} \frac{1}{n} \int_{\Omega_{n,r}} |\ddot{K}_{n,r,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} &= (2\pi)^{-\frac{1}{2}} \int_{\tilde{\Omega}_{n,r}} \sqrt{n}|nt_r^2 - 1| e^{-\frac{nt_r^2}{2}} dt_r \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \\ &\leq \frac{1}{\pi\sqrt{n}} \left(\left[nt_r e^{-\frac{nt_r^2}{2}} \right] \Big|_{t_r=1/\sqrt{n}}^{2\sqrt{\frac{\log n}{n}}} + \left[nt_r e^{-\frac{nt_r^2}{2}} \right] \Big|_{t_r=1/\sqrt{n}}^0 \right) = O(1), \end{aligned}$$

where the inequality follows from $0 \leq \int \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq 1$.

$$\begin{aligned} \text{For (iii), by definition, } \int_{\Omega_{n,r}} \ddot{K}_{n,r,r}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} &= \int_{\tilde{\Omega}_{n,r}} \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} (n^2 t_r^2 - n) e^{-\frac{n\|\mathbf{t}\|_2^2}{2}} d\mathbf{t} \\ &= 2n^{\frac{3}{2}} t_r e^{-\frac{nt_r^2}{2}} \Big|_{t_r=0}^{2\sqrt{\frac{\log n}{n}}} \times \int_{\tilde{\Omega}_{n,r}} \prod_{i \neq r} d\Phi(\sqrt{n}t_i) = o\left(\frac{1}{\sqrt{n}}\right), \text{ where the last equality} \\ &\text{follows from } 0 \leq \int \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq 1. \end{aligned}$$

For (iv), by definition and applying integration by parts twice,

$$\begin{aligned} &\int_{\Omega_{n,r}} \frac{1}{2} (\mathbf{t} - \boldsymbol{\theta})' \mathbf{A} (\mathbf{t} - \boldsymbol{\theta}) \ddot{K}_{n,r,r}(\mathbf{t}, \boldsymbol{\theta}) d\mathbf{t} \\ &= \int_{\tilde{\Omega}_{n,r}} \frac{1}{2} \mathbf{t}' \mathbf{A} \mathbf{t} (2\pi)^{-\frac{1}{2}} \sqrt{n} d(-nt_r e^{-\frac{nt_r^2}{2}}) \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \\ &= o(n^{-1}) + \int_{\tilde{\Omega}_{n,r}} \mathbf{t}' \mathbf{A} \mathbf{e}_r (2\pi)^{-\frac{1}{2}} \sqrt{n} d(-e^{-\frac{nt_r^2}{2}}) \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \\ &= \mathbf{A}_{r,r} + o(n^{-1}), \end{aligned}$$

where $\mathbf{e}_r' = (0, \dots, 0, 1, 0, \dots, 0)$ with r^{th} entry being 1, and the last equality follows from the Gaussian tail probability.

For (v), we know, by definition, $\int_{\Omega_{n,r}^C} |\dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} = \int_{\tilde{\Omega}_{n,r}^C} |\dot{K}_{n,r}(\mathbf{t}, \mathbf{0})| d\mathbf{t}$.

By symmetry,

$$\begin{aligned} \int_{\tilde{\Omega}_{n,r}^{(1)}} |\dot{K}_{n,r}(\mathbf{t}, \mathbf{0})| d\mathbf{t} &= \frac{2}{\sqrt{2\pi}} \int_{t_r \geq 0, \tilde{\Omega}_{n,r}^{(1)}} n^{\frac{1}{2}} d(e^{-\frac{nt_r^2}{2}}) \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \\ &\leq \frac{\sqrt{n}}{n^2} \times 1 = O(n^{-\frac{3}{2}}), \end{aligned}$$

where the inequality follows from $0 \leq \int \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq 1$. Similarly,

$$\int_{\tilde{\Omega}_{n,r}^{(2)}} |\ddot{K}_{n,r,r}(\mathbf{t}, \mathbf{0})| d\mathbf{t} = \frac{2\sqrt{n}}{\sqrt{2\pi}} \int_{t_r \geq 0, \tilde{\Omega}_{n,r}^{(2)}} d(e^{-\frac{nt_r^2}{2}}) \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq \frac{\sqrt{n}}{n^2} = O(n^{-\frac{3}{2}}).$$

For (vi), by definition,

$$\begin{aligned} \frac{1}{\sqrt{n}} \int_{\Omega_{n,r}} |\dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} &= \int_{\tilde{\Omega}_{n,r}} (2\pi)^{-\frac{d}{2}} n^{\frac{d-1}{2}} n |t_r| e^{-\frac{n\|\mathbf{t}\|_2^2}{2}} d\mathbf{t} \\ &= (2\pi)^{-\frac{1}{2}} 2 \int_{t_r \geq 0, \tilde{\Omega}_{n,r}} d(e^{-\frac{nt_r^2}{2}}) \prod_{i \neq r} d\Phi(\sqrt{n}t_i) = O(1), \end{aligned}$$

where the second equality is due to symmetry and the third equality follows from $0 \leq \int \prod_{i \neq r} d\Phi(\sqrt{n}t_i) \leq 1$.

To prove (vii), without loss of generality, we assume $0 \leq G(\mathbf{t}, \boldsymbol{\omega}) \leq 1$. We denote $\Omega_{n,r}^a$ as $\Omega_{n,r}$ and $\Omega_{n,r}^b$ its complement. Then,

$$\begin{aligned} \int_{\Omega_{n,r}^k \times \Omega_{n,s}^l} G(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} \\ \leq \int_{\Omega_{n,r}^k} |\dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta})| d\mathbf{t} \times \int_{\Omega_{n,s}^l} |\dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta})| d\boldsymbol{\omega} \\ = \int_{\tilde{\Omega}_{n,r}^k} |\dot{K}_{n,r}(\mathbf{t}, \mathbf{0})| d\mathbf{t} \times \int_{\tilde{\Omega}_{n,s}^l} |\dot{K}_{n,s}(\boldsymbol{\omega}, \mathbf{0})| d\boldsymbol{\omega} \end{aligned}$$

where k and l are chosen from $\{a, b\}$. Then by (v) and (vi),

$$\begin{aligned} \int G(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} \\ = \int_{\Omega_{n,r} \times \Omega_{n,s}} G(\mathbf{t}, \boldsymbol{\omega}) \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}) \dot{K}_{n,s}(\boldsymbol{\omega}, \boldsymbol{\theta}) d\mathbf{t} d\boldsymbol{\omega} + o(n^{-\frac{1}{2}}). \end{aligned}$$

□

LEMMA 3. *Uniformly over $(\mathbf{t}, \boldsymbol{\omega})$ such that $\|\mathbf{t} - \boldsymbol{\theta}_0\| = o(1)$ and $\|\boldsymbol{\omega} - \boldsymbol{\theta}_0\| = o(1)$, we have*

$$(A.1) \quad \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \tau(\mathbf{u}_i, \mathbf{t})}{\partial \theta_r} \tau(\mathbf{u}_i, \boldsymbol{\omega}) \right] = E \left[\frac{\partial \tau(\mathbf{u}, \boldsymbol{\theta}_0)}{\partial \theta_r} \tau(\mathbf{u}, \boldsymbol{\theta}_0) \right] + o_p(1),$$

$$(A.2) \quad \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \tau(\mathbf{u}_i, \mathbf{t})}{\partial \theta_r} \frac{\partial \tau(\mathbf{u}_i, \boldsymbol{\omega})}{\partial \theta_s} \right] = E \left[\frac{\partial \tau(\mathbf{u}, \boldsymbol{\theta}_0)}{\partial \theta_r} \frac{\partial \tau(\mathbf{u}, \boldsymbol{\theta}_0)}{\partial \theta_s} \right] + o_p(1).$$

PROOF. We sketch the main steps of the proof below.

First of all, observe that

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \tau(\mathbf{u}_i, \mathbf{t})}{\partial \theta_r} \tau(\mathbf{u}_i, \boldsymbol{\omega}) \right] - \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \tau(\mathbf{u}_i, \boldsymbol{\theta}_0)}{\partial \theta_r} \tau(\mathbf{u}_i, \boldsymbol{\omega}) \right] \right| \\ \leq \frac{1}{n} \sum_{i=1}^n \left[\left| \frac{\partial \tau(\mathbf{u}_i, \mathbf{t})}{\partial \theta_r} - \frac{\partial \tau(\mathbf{u}_i, \boldsymbol{\theta}_0)}{\partial \theta_r} \right| \times |\tau(\mathbf{u}_i, \boldsymbol{\omega})| \right] \\ \leq \frac{1}{n} \sum_{i=1}^n M_2(\mathbf{u}_i) \times |\mathbf{t} - \boldsymbol{\theta}_0|, \end{aligned}$$

where $M_2(\mathbf{u})$ is an integrable function. The last inequality is due to Assumption 3 and $|\tau(\mathbf{u}, \boldsymbol{\theta})| \leq 1$.

Since $M_2(\mathbf{u})$ is integrable, by the law of large numbers, the left hand side of above inequality is thus $o_p(1)$. By a similar argument, we can show that

$$\frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \tau(\mathbf{u}_i, \mathbf{t})}{\partial \theta_r} \tau(\mathbf{u}_i, \boldsymbol{\omega}) \right] = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \tau(\mathbf{u}_i, \boldsymbol{\theta}_0)}{\partial \theta_r} \tau(\mathbf{u}_i, \boldsymbol{\theta}_0) \right] + o_p(1).$$

By the law of large numbers, we get (A.1). The proof of (A.2) is similar. \square

LEMMA 4. *Let $\hat{\mathbf{A}}_n$ and $\hat{\mathbf{V}}_n$ be the same as those in (7). Then, for $1 \leq r, s \leq d$, we have*

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = o(1), \boldsymbol{\Sigma} \in \mathcal{N}(\mathbf{D}_0)} \left| \frac{\partial}{\partial \boldsymbol{\Sigma}} [\hat{\mathbf{A}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})]_{r,s} \right| = o_p(1),$$

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = o(1), \boldsymbol{\Sigma} \in \mathcal{N}(\mathbf{D}_0)} \left| \frac{\partial}{\partial \boldsymbol{\Sigma}} [\hat{\mathbf{V}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})]_{r,s} \right| = o_p(1),$$

where $\mathcal{N}(\mathbf{D}_0)$ is a small neighborhood of \mathbf{D}_0 and $\boldsymbol{\Sigma}$ is a positive definite matrix.

PROOF. We now extend the definition of kernels in Lemma 2 for any covariance matrix Σ as follows,

$$K_n(\mathbf{t}, \boldsymbol{\theta}, \Sigma) := \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} |\Sigma|^{-1/2} \exp\left(-\frac{n}{2}(\mathbf{t} - \boldsymbol{\theta})' \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta})\right),$$

where $|\Sigma|$ is the determinant of Σ . Then the first and second derivatives of K_n with respect to $\boldsymbol{\theta}$ become

$$\begin{aligned} \dot{K}_{n,r}(\mathbf{t}, \boldsymbol{\theta}, \Sigma) &:= \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} n \mathbf{e}_r' \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta}) e^{-\frac{n}{2}(\mathbf{t} - \boldsymbol{\theta})' \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta})}, \\ \ddot{K}_{n,r,s}(\mathbf{t}, \boldsymbol{\theta}, \Sigma) &:= \left(\frac{2\pi}{n}\right)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} n \mathbf{e}_r' \left[n \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})' \Sigma^{-1} - \Sigma^{-1} \right] \mathbf{e}_s \\ &\quad \times e^{-\frac{n}{2}(\mathbf{t} - \boldsymbol{\theta})' \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta})}. \end{aligned}$$

We partition \mathbb{R}^d into $\Omega_{n,r}$ and its complement $\Omega_{n,r}^c$, where $\Omega_{n,r} := \left\{ \mathbf{t} : (\mathbf{t} - \boldsymbol{\theta})' \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta}) < 6d \log n/n \right\}$. Furthermore, We define $\tilde{\Omega}_{n,r} := \left\{ \mathbf{t} : \mathbf{t}' \Sigma^{-1} \mathbf{t} < 6d \log n/n \right\}$.

Note that $(\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})' e^{-\frac{1}{2}(\mathbf{t} - \boldsymbol{\theta})' \Sigma^{-1}(\mathbf{t} - \boldsymbol{\theta})}$ is bounded for $\Sigma \in \mathcal{N}(D_0)$. Similar to the proofs of Theorem 2 and Lemma 2, we can get

$$\begin{aligned} \frac{\partial[\hat{\mathbf{A}}_n(\boldsymbol{\theta}, \Sigma)]_{r,s}}{\partial \Sigma} &= [\mathbf{A}(\boldsymbol{\theta}_0)]_{r,s} \int_{\tilde{\Omega}_{n,r}} \frac{\partial K_n(\mathbf{t}, \mathbf{0}, \Sigma)}{\partial \Sigma} d\mathbf{t} + o_p(1), \\ \frac{\partial[\hat{\mathbf{V}}_n(\boldsymbol{\theta}, \Sigma)]_{r,s}}{\partial \Sigma} &= [\mathbf{V}(\boldsymbol{\theta}_0)]_{r,s} \int_{\tilde{\Omega}_{n,r}} \frac{\partial K_n(\mathbf{t}, \mathbf{0}, \Sigma)}{\partial \Sigma} d\mathbf{t} + o_p(1), \end{aligned}$$

uniformly over $(\boldsymbol{\theta}, \Sigma)$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = o(1)$ and $\|\Sigma - \mathbf{D}_0\| = o(1)$.

Likewise, we have $\int_{\tilde{\Omega}_{n,r}^c} \frac{\partial K_n(\mathbf{t}, \mathbf{0}, \Sigma)}{\partial \Sigma} d\mathbf{t} = o(1)$, which, combined with $\int \frac{\partial K_n(\mathbf{t}, \mathbf{0}, \Sigma)}{\partial \Sigma} d\mathbf{t} = 0$, implies $\int_{\tilde{\Omega}_{n,r}} \frac{\partial K_n(\mathbf{t}, \mathbf{0}, \Sigma)}{\partial \Sigma} d\mathbf{t} = o(1)$. This completes the proof. \square

COROLLARY 1. For $1 \leq r, s \leq d$, we have

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = o(1), \Sigma \in \mathcal{N}(D_0)} \left| \frac{\partial}{\partial \Sigma} [\hat{\mathbf{A}}_n(\boldsymbol{\theta}, \Sigma)^{-1}]_{r,s} \right| = o_p(1).$$

PROOF. First, by Theorem 2, Lemma 4 and the mean value theorem, we can show that $[\hat{\mathbf{A}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})]_{r,s} = [\hat{\mathbf{A}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma}) - \hat{\mathbf{A}}_n(\boldsymbol{\theta}, \mathbf{D}_0) + \hat{\mathbf{A}}_n(\boldsymbol{\theta}, \mathbf{D}_0)]_{r,s} = [\mathbf{A}(\boldsymbol{\theta}_0)]_{r,s} + o_p(1)$. By matrix differentiation, $d\mathbf{A}^{-1} = -\mathbf{A}^{-1}(d\mathbf{A})\mathbf{A}^{-1}$. Thus $\hat{\mathbf{A}}_n^{-1} - \mathbf{A}_0^{-1} = -\mathbf{A}_0^{-1}(\hat{\mathbf{A}}_n - \mathbf{A}_0)\mathbf{A}_0^{-1} + o(\|\hat{\mathbf{A}}_n - \mathbf{A}_0\|_1)$, where $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\theta}_0)$. The rest of the proof is straightforward and thus omitted. \square

LEMMA 5. For $1 \leq r, s \leq d$, we have

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = o(1), \boldsymbol{\Sigma} \in \mathcal{N}(\mathbf{D}_0)} \left| \frac{\partial}{\partial \boldsymbol{\Sigma}} [\hat{\mathbf{D}}_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})]_{r,s} \right| = o_p(1).$$

PROOF. The result follows immediate from Lemma 4 and Corollary 1. \square

APPENDIX B: A SUFFICIENT CONDITION FOR ASSUMPTION 3

Suppose f is the joint density for (\mathbf{X}, Y) and $f(\cdot | \mathbf{r}, s)$ is the conditional density of $X^{(2)}$ given $\mathbf{X}^{(1)} = \mathbf{r}$ and $Y = s$. Suppose $g(\cdot | s, \boldsymbol{\theta})$ is the conditional density of $\mathbf{X}'\boldsymbol{\beta}(\boldsymbol{\theta})$ given $Y = s$ and $g(\cdot | \mathbf{r}, s, \boldsymbol{\theta})$ is the conditional density of $\mathbf{X}'\boldsymbol{\beta}(\boldsymbol{\theta})$ given $\mathbf{X}^{(1)} = \mathbf{r}$ and $Y = s$. By change of variable, $g(\mathbf{t} | \mathbf{r}, s, \boldsymbol{\theta}) = f(\mathbf{t} - \mathbf{r}'\boldsymbol{\theta} | \mathbf{r}, s)$. Therefore,

$$g(\mathbf{t} | s, \boldsymbol{\theta}) = \int g(\mathbf{t} | \mathbf{r}, s, \boldsymbol{\theta}) G_{\mathbf{X}^{(1)} | s}(d\mathbf{r}) = \int f(\mathbf{t} - \mathbf{r}'\boldsymbol{\theta} | \mathbf{r}, s) G_{\mathbf{X}^{(1)} | s}(d\mathbf{r}),$$

where $G_{\mathbf{X}^{(1)} | s}$ is the conditional distribution of $\mathbf{X}^{(1)}$ given $Y = s$. We also observe that,

$$\tau(z, \boldsymbol{\theta}) = \int_{-\infty}^{x'\boldsymbol{\beta}(\boldsymbol{\theta})} \int_{-\infty}^y g(\mathbf{t} | s, \boldsymbol{\theta}) G_Y(ds) d\mathbf{t} + \int_{x'\boldsymbol{\beta}(\boldsymbol{\theta})}^{\infty} \int_y^{\infty} g(\mathbf{t} | s, \boldsymbol{\theta}) G_Y(ds) d\mathbf{t},$$

where G_Y is the marginal distribution of Y . Therefore if the conditional density $f_{\mathbf{X}^{(2)} | \mathbf{X}^{(1)}, Y}(\cdot | \mathbf{r}, s)$ has bounded derivatives up to order three for each (\mathbf{r}, s) in the support of space $\mathbf{X}^{(1)} \otimes Y$, it is not difficult to show that Assumption 3 is satisfied. The sufficient condition can be easily verified in certain common situations such as when the conditional density $f_{\mathbf{X}^{(2)} | \mathbf{X}^{(1)}, Y}$ is normal.

REFERENCES

- [1] BENNETT, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.*, **2**, 273-277.
- [2] BOX, G. E. P. AND COX, D. R. (1964). An analysis of transformations. *J. R. Statist. Soc. B*, **26**, 211-252.
- [3] BROWN, B. M. AND WANG, Y. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, **92**, 149-158.
- [4] BROWN, B. M. AND WANG, Y. (2007). Induced smoothing for rank regression with censored survival times. *Statist. Med.*, **26**, 828-836.

- [5] CHEN, S. (2002). Rank estimation of transformation models. *Econometrica*, **70**, 1683-1697.
- [6] COX, D. R. (1972). Regression models and life tables. *J. R. Statist. Soc. B*, **34**, 187-220.
- [7] COX, D. R. AND OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- [8] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist*, **7**, 1-26.
- [9] GØRGENSEN, T. AND HOROWITZ, J. L. (1999). Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *J. Econometrics*, **90**, 155-191.
- [10] HAN, A. K. (1987). Non-parametric analysis of a generalized regression model. *J. Econometrics*, **35**, 303-316.
- [11] HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica*, **46**, 1251-1271.
- [12] JIN, Z., YING, Z. AND WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika*, **88**, 381-390.
- [13] KALBFLEISCH, J. D. AND PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data (2nd Ed.)*. New York: Wiley.
- [14] KHAN, S. AND TAMER, E. (2007). Partial rank estimation of duration models with general forms of censoring. *J. Econometrics*, **136**, 251-280.
- [15] MADDALA, G. S. (1983). Limited dependent and qualitative variables in econometrics. *Econometric Society Monograph No.3*. Cambridge: Cambridge University Press.
- [16] MCFADDEN, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of Econometrics, Vol.2*, 1395-1457, Amsterdam: Elsevier.
- [17] NOLAN, D. AND POLLARD, D. (1987). U-processes: rates of convergence. *Ann. Statist*, **15**, 780-799.
- [18] POWELL, J. L. (1984). Least absolute deviations estimation for the censored regression model. *J. Econometrics*, **25**, 303-325.
- [19] SHERMAN, R. P. (1993). The limit distribution of the maximum rank correlation estimator. *Econometrica*, **61**, 123-137.
- [20] SHERMAN, R. P. (1994a). U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory*, **10**, 372-395.
- [21] SHERMAN, R. P. (1994b). Maximum inequalities for degenerate U-processes with applications to maximization estimators. *Ann. Statist.*, **22**, 439-459.
- [22] TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24-26.
- [23] VAN DER VAART, A.W. (1998). *Asymptotic Statistics.*, New York: Cambridge University Press.

J. ZHANG
 Z. YING
 DEPARTMENT OF STATISTICS,
 COLUMBIA UNIVERSITY,
 2690 BROADWAY,
 NEW YORK, NY10027, U.S.A.
 E-MAIL: jj2299@columbia.edu
 E-MAIL: zying@stat.columbia.edu

Z. JIN
 DEPARTMENT OF BIOSTATISTICS,
 MAILMAN SCHOOL OF PUBLIC HEALTH,
 COLUMBIA UNIVERSITY,
 NEW YORK, NY10032, U.S.A.
 E-MAIL: zjin@biostat.columbia.edu

Y. SHAO
DIVISION OF BIostatISTICS,
SCHOOL OF MEDICINE,
NEW YORK UNIVERSITY,
NEW YORK, NY10016, U.S.A.
E-MAIL: yongzhao.shao@nyumc.org